Biometrics 59, 1–24 DOI: 000
December 2013

Multilevel Quantile Function Modeling with Application to Birth Outcomes

Luke B. Smith[1,*], Montserrat Fuentes[1], Brian J. Reich [1], Amy H. Herring[2], and Peter H. Langlois[3]

[1]Department of Statistics, North Carolina State University, Raleigh, North Carolina, 27695-8203, U.S.A.

[2]Department of Biostatistics, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina, 27599-7420, U.S.A.

[3]Texas Department of State Health Services, Austin, Texas 78714-9347, U.S.A.

email: luke_smith@ncsu.edu

Summary: Infants born preterm or small for gestational age have elevated rates of morbidity and mortality. Using birth certificate records in Texas from 2002-2004 and Environmental Protection Agency air pollution estimates, we relate the quantile functions of birth weight and gestational age to ozone exposure and multiple predictors, including parental age, race, and education level. We introduce a semi-parametric Bayesian quantile approach that models the full quantile function rather than just a few quantile levels. Our multilevel quantile function model establishes relationships between birth weight and the predictors separately for each week of gestational age and between gestational age and the predictors separately across Texas Public Health Regions. We permit these relationships to vary nonlinearly across gestational age, spatial domain and quantile level and we unite them in a hierarchical model via a basis expansion on the regression coefficients that preserves interpretability. Very low birth weight is a primary concern, so we leverage extreme value theory to supplement our model in the tail of the distribution. Gestational ages are rounded into weekly values, so we present methodology for modeling quantile functions of discrete response data. In a simulation study we show that pooling information across gestational age and quantile level substantially reduces MSE of predictor effects relative to standard frequentist quantile regression. We find that ozone is negatively associated with the lower tail of gestational age in south Texas and across the distribution of birth weight for high gestational ages. Our methods are available in the R package **BSquare**.

Key words: Birth weight; Discrete; Extremes; Gestational Age; Graphics processing units; Ozone; Quantile.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Infants who are born preterm (gestational period less than 37 weeks) or small for gestational age (below the 10th percentile of birth weight after controlling for gestational age) have elevated rates of morbidity and mortality (Honein et al., 2009; Pulver et al., 2009; Garite et al., 2004). Reasons for these associations include poorly functioning organs, reduced metabolism, insulin resistance, and increased susceptibility to adverse environmental events later in life (Barker, 2006). Infants who are both preterm and small for gestational age (SGA) are at higher mortality risk than infants with either condition singly (Katz et al., 2013). Narchi et al. (2010) found that adjusting the conditional distribution of birth weight for biological variables better identified at risk infants.

Our first scientific objective is to better define the conditional distributions of gestational age and birth weight by incorporating personal characteristics and environmental factors. We use information from Texas birth certificate records, including maternal parity, sex of the infant, parental education level, parental age and race. In a paper with similar aims, Gardosi et al. (1995) used stepwise regression to define the conditional percentiles. We want to understand the relationship between the predictors and the tails of these variables, so we model the conditional quantile functions of the birth outcomes. In a literature review Šrám et al. (2005) argued that the relationships between air pollution and gestational age and intrauterine growth warrant further analysis. Our second scientific objective is to investigate the effect of maternal exposure to tropospheric ozone, one of the criteria pollutants regulated under the Environmental Protection Agency's Clean Air Act, on SGA and preterm birth (PTB).

Classical frequentist (Koenker, 2005) and Bayesian (Yu and Moyeed, 2001) quantile regression models a conditional quantile rather than the conditional mean as a function of predictors. This enables inference of noncentral parts of the distribution, makes fewer

assumptions, and is more robust to outliers than mean regression. One limitation with these approaches is that fits at multiple levels can produce "crossing quantiles," where for some values of the predictors the quantile function is decreasing in quantile level. Modeling multiple quantile levels through constraints on the coefficients ensures monotonicity of the quantile function, as in (Bondell et al., 2010) and the references therein.

The aforementioned approaches model a finite number of quantile levels and do not share information across quantile level. In applications where we expect inference at proximate quantile levels to be similar, it is useful to encourage communication across the distribution. Specifying the full quantile function, which entails separate parameter effects at an uncountable number of quantile levels, fosters this all-encompassing approach. Recent examples of quantile function modeling include Reich et al. (2011), who investigated the effects of temperature on tropospheric ozone using Bernstein polynomials, and Tokdar and Kadane (2011), who analyzed birth weights using stochastic integrals. Reich and Smith (2013) extended quantile function methodology to censored data.

We faced three methodological hurdles in our application. PTB and low birth weight are closely related, but distinct, concerns. Researchers prefer to define SGA infants to isolate effects on birth weight from those on gestational age, so it is important to allow the relationship between birth weight and the predictors to vary by gestational age. While multilevel regression models are well-suited for jointly modeling a collection of distributions, standard hierarchical models assume the predictors affect only the conditional mean of the response. Second, considerable interest lies in the tails (particularly in very premature, SGA or large-for-gestational age births), so it is important to enable the tails of these distributions to be affected differentially by the predictors relative to the center of the distribution. Estimation of parameter effects at very low or very high quantiles is generally the purview of extreme value analysis. Multiple conditional extremal methods exist in the literature. Wang

and Tsai (2009) modeled the tail index, which determines the thickness of the tails, through a linear log link function of the parameters. Wang et al. (2012) quantile regressed in the shallow tails and extrapolated the results into the deep tails for thickly-tailed data. Our application requires inference across the distribution, so we follow the approaches of (Zhou et al., 2012) and (Reich et al., 2011), who modeled the middle of the distribution semiparametrically and fit a parametric form above a threshold. In these applications either zero (Zhou et al., 2012) or one (Reich et al., 2011) covariate affected the distribution above the threshold. Our final methodological challenge is modeling a discrete response. The gestational age measurements have been rounded into values of $\{25, 26, ...42\}$ weeks. Canonical discrete regression models make restrictive assumptions about the relationship between the response and the predictors. Dichotomizing the response by PTB restricts inference to the cutpoint between 36 and 37 weeks. Previous approaches in the literature (Machado and Silva, 2005) modeled one quantile by adding random noise to compel the response to behave continuously.

The primary contribution of this paper is to introduce a class of multilevel quantile function models that overcomes these methodological challenges. The distribution of birth weight transitions smoothly across gestational age, as shown in Figure 1.

[Figure 1 about here.]

We exploit this smoothness by jointly modeling birth weight as a dependent collection of distributions ordered by gestational age. The horizontal lines in Figure 1 represent the thresholds for low birth weight (LBW), defined as 2500 grams, and very low birth weight (VLBW), defined as less than 1500 grams (Rogers and Dunlop, 2006). Most infants born at 25 weeks of gestational age are classified as VLBW, while almost no infants born at 39 weeks and greater are VLBW, so it is imperative to control for gestational age when examining fetal-restricted growth. Our multilevel approach obviates the choice between the high flexibility and low power associated with separate fits across gestational ages and the

high power and inflexibility derived from one fit for all gestational ages. We illustrate another example of our multilevel class by spatially correlating separate distributions of gestational age for each of the eleven Texas Public Health Regions, which are shown in Figure 2.

[Figure 2 about here.]

In both cases we cohere the individual models via Gaussian process priors on the regression coefficients. This class fits separate regression parameters for different gestational ages/spatial regions, quantile levels and predictors, creating a rich environment for parameter estimation.

Our second methodological contribution is a synthesis of quantile function modeling and conditional extreme value analysis. We adopt a semiparametric approach that models the middle of the distribution as a linear combination of basis functions and parametrically fits the tails of the distribution via a smooth transition across the semiparametric/parametric threshold. We implement a quantile model that is parameterized to detect changes due to the predictors in the lower tails and can be used to recover the conditional probabilities of PTB and SGA. This enhances tail flexibility, ensuring inference on the quantile levels of interest is not perturbed by a few outliers in the tails.

Our final methodological contribution is to extend the quantile function approach to accommodate discrete data. We model the discretized response as a censored realization of a latent continuous process. By modeling the full quantile function we can estimate predictor effects in a computationally stable manner. To our knowledge this is the first quantile function model that can handle discrete responses.

The paper is structured as follows. In Section 2 we describe the hierarchical quantile model. In Section 3 we present the results of a simulation study that explores our three methodological innovations. In Section 4 we analyze the birth outcomes and we conclude in Section 5.

2. Quantile Function Modeling

Denote Y_i as the response (either birth weight or gestational age as described below) and X_i as the vector of length p containing personal characteristics, environmental variables and intercept of infant i. We can define the model for Y_i by the conditional distribution function $F(y;X) = P(Y \leqslant y;X)$ or the density $f(y;X) = \frac{\mathrm{d}}{\mathrm{d}y}F(y;X)$. Alternatively we can specify Y by its conditional quantile function $Q(\tau;X)$ where $Q(\tau;X) = F^{-1}(\tau;X) = \inf\{y : F(y;X) \geqslant \tau\}$. The value $\tau \in [0,1]$ is known as the quantile level and the quantile function is nondecreasing in the quantile level. Birth weight has been previously modeled in the quantile regression (Koenker and Hallock, 2001; Burgette and Reiter, 2012; Tokdar and Kadane, 2011), density estimation (Dunson et al., 2008) and spatial (Kammann and Wand, 2003) settings. In this paper we borrow from all of these domains.

We begin with the class of bounded distributions, where there exists real numbers a and b such that for all X, $a < Q(0;X) < Q(1;X) < b$. For now we assume the density of Y is absolutely continuous with respect to Lebesgue measure, implying a unique quantile function that is increasing in quantile level. We describe extensions to cases of unbounded distributions and discrete response in Sections 2.3 and 2.4 respectively.

2.1 *Individual Quantile Function*

The most flexible method would allow the predictors to nonlinearly affect the quantile function. This approach is promising for prediction, but the nonlinearity of the predictor effects makes inference challenging, so we model the parameter effects for each predictor to be linear at each quantile level. Our quantile function is

$$Q(\tau;X_i) = \sum_{j=1}^{P} X_{ij}\beta_j(\tau) = \sum_{j=1}^{P} X_{ij} \sum_{m=1}^{M} I_m(\tau)\theta_{mj} \tag{1}$$

where $I_1(\tau) \equiv 1$, $I_m(\tau)$ is an increasing basis function in τ for $m > 1$ and θ_{mj} are the regression parameters.

The derivative of the quantile function with respect to X_j is $\beta_j(\tau) = \sum_{m=1}^{M} I_m(\tau)\theta_{mj}$ and

is free of all elements of X for any j. A one-unit increase in X_j is associated with a $\beta_j(\tau)$ increase in the τ^{th} quantile of the response. Mean regression is a special case of this model where $\theta_{mj} \equiv 0$ for $j > 1$ and $m > 1$ and the effect on the mean from the j^{th} predictor is θ_{j1}.

The quantile function is increasing in τ if the derivative of the quantile function with respect to τ is positive. This derivative $q(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}Q(\tau)$ is known as the sparsity function (Tukey, 1965; Parzen, 1979) and is the reciprocal of the density for any valid differentiable quantile function (note that $F(Q(\tau)) = \tau$ and differentiate both sides with respect to τ.) Therefore, we can start with any valid quantile function and find its likelihood, as in Tokdar and Kadane (2011). In our model the sparsity function is

$$q(\tau; X_i) = \sum_{m=1}^{M}\sum_{j=1}^{P} X_{ij}B_m(\tau)\theta_{mj} \tag{2}$$

where $B_m(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}I_m(\tau)$. This produces the likelihood

$$L(\theta, Y) = \prod_{i=1}^{N} q(\tau_i; X_i)^{-1}$$

where $\tau_i = Q^{-1}(Y_i)$.

Similar to Reich and Smith (2013), we map all predictors into $[-1, 1]$ and need $\theta_{m1} > \sum_{j=2}^{P}|\theta_{mj}|$ for m > 1 to ensure a valid quantile process. Let $\tau \in [0, 1]$ be arbitrary. Let $X_j = 1$ if the j^{th} predictor has a negative effect at τ and $X_j = -1$ otherwise. This is the "worst case" combination of the predictors that minimizes the sparsity function. Then the sparsity function $q(\tau; X_i) = \sum_{m=1}^{M} B_m(\tau)\sum_{j=1}^{P} X_{ij}\theta_{mj} \geqslant \sum_{m=1}^{M} B_m(\tau)(\theta_{m1} - \sum_{j=2}^{P}; \theta_{mj}|) > 0$ because $\theta_{m1} > \sum_{j=2}^{P}|\theta_{mj}|$. The derivative of the quantile function is positive for all τ and therefore is increasing in τ. We put a Gaussian process prior on latent coefficients $\theta_m^\star = (\theta_{m1}, ..., \theta_{mP})$. We let $\theta_{mj} = \theta_{mj}^\star$ for all j if the monotonicity constraints are satisfied for the m^{th} basis function and set

$$\theta_{mj} = \begin{cases} 0.001 & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

if $\theta^\star$ is outside of the constraint space.

The distribution of the response is bounded, so the least restrictive case is to assume that for any value of the predictors the quantile function of birth weight is in the space of bounded monotonically increasing functions. In this paper we model the projection of the quantile function onto the space of cubic integrated M-splines, known as I-splines of degree 3 (Ramsay, 1988). Let $t = \{t_0, ..., t_K\}$ be an ordered sequence of knots whose minimum value is $t_0 = 0$ and maximum value is $t_K = 1$. In the k^{th} interval between knots t_k and t_{k+1} I-splines of degree 3 are locally cubic polynomials, and at the knots I-splines are continuous and first-order differentiable. As the number of knots increases the space of polynomial splines converges to the space of continuous functions (Schumaker, 1981). Additionally, the space of cubic monotonic splines converges to the space of continuous monotonic functions almost as quickly as unconstrained cubic splines (DeVore, 1977). We supplement the cubic spline basis with a constant basis function.

2.2 *Multilevel Quantile Model*

In the birth weight application we individually relate the distribution of birth weight to the predictors for each of the $G = 18$ gestational ages. We model these distributions collectively to gain power by borrowing information across gestational age while reducing the loss of model richness. To spur communication across gestational ages the elements of $\theta^\star$ corresponding to the basis function m and predictor j are collectively modeled as a multivariate normal distribution with common mean μ_{mj} and precision γ_{mj}. This lower-rank representation of the individual fits reduces the variance of the regression coefficients. The quantile function at one gestational age in our global fit is of the form

$$Q(\tau; X_i, GA = g) = \sum_{m=1}^{M} \sum_{j=1}^{P} X_{ij} I_m(\tau) \theta_{mjg}. \tag{3}$$

As in the individual quantile function case, we let $\theta_{mj.}$ be the observed realization of the latent $\theta^\star_{mj.} \sim \text{N}(\mu_{mj}, \gamma_{mj}^{-1}\Sigma_{mj})$ that satisfies the monotonicity constraints at each gestational

age g. This hierarchical framework provides the potential for sharing information. If the collection of distributions are similar the gain in power can be substantial.

The correlation Σ_{mj} is used to capture patterns in the basis functions after shrinking them to a common mean. In our applications Σ_{mj} is used to smooth gestational age quantile functions over spatial health regions and birth weight quantile functions over gestational age. For gestational age we are interested in examining the distribution regionally. To capture spatial dependence in the quantile function we fit an exponential spatial correlation matrix, where $\Sigma_{mj}[u, v] = \exp\{-\phi_{mj} d(u, v)\}$ and $d(u, v)$ is the distance between the centroids of public health regions u and v. We assign a uniform prior to the range parameter with minimum 0 and maximum equal to half the maximum distance between public health region centroids. This approach enables us to examine large scale spatial patterns in the relationship between gestational age and the predictors. For birth weight we anticipate dependence across gestational age, so we impose an autoregressive order 1 correlation matrix of the form $\Sigma_{mj}[u, v] = \exp\{-\rho_{mj}|u - v|\}$ with correlation parameter ρ_{mj}.

2.3 *Tail Modeling*

For extreme quantile levels, we employ a hybrid approach of quantile modeling and extremal analysis. While there are lower bounds for the birth outcomes, modeling the quantile function of birth weight using bounded basis functions is restrictive. If any observations fall below $Q(0; X_i)$ or above $Q(1; X_i)$ the likelihood is zero, so bounded basis functions result in a discontinuous likelihood with respect to the parameters. Both expert opinion (Wilcox, 2001) and exploratory analysis suggest the density of birth weight has thick tails, which are best modeled using basis functions that decay slowly in the tails. We model the middle of the distribution with a linear combination of nonparametric basis functions and the tails by a member of the generalized Pareto distribution (GPD) family. The GPD is the canonical method in extremal inference for fitting exceedances over a threshold. Most standard statis-

tical distributions have tail behavior conforming to the GPD (Coles, 2001). The type of tail behavior is determined by the shape parameter ξ and corresponds to bounded distributions ($\xi < 0$), light tails ($\xi = 0$) and heavy tails ($\xi > 0$). These densities are respectively of the form

$$\zeta_Z(z;\sigma,\xi) = \begin{cases} \sigma^{-1}(1+\frac{\xi z}{\sigma})^{-1/\xi-1}\mathbb{1}_{0<z<-\sigma/\xi} & \xi < 0 \\ \sigma^{-1}\exp(\frac{-z}{\sigma})\mathbb{1}_{z>0} & \xi = 0 \\ \sigma^{-1}(1+\frac{\xi z}{\sigma})^{-1/\xi-1}\mathbb{1}_{z>0} & \xi > 0 \end{cases}$$

where $\mathbb{1}$ is the indicator function. We set a lower threshold at quantile level τ_L and fit a GPD if $Y_i < Q(\tau_L; X)$. We do the same for upper threshold τ_H. This methodology was developed in Zhou et al. (2012) for continuous spatial data without predictors. We extend this approach to incorporate covariates and discrete data.

The scale parameters (σ_L, σ_H) are the density of the Pareto distribution evaluated at the thresholds $(Q(\tau_L; X), Q(\tau_H; X))$, as can be seen by plugging $z = 0$ into the density. Assuming a continuous density specifies the scale parameters as

$$\sigma_L|X = \tau_L * q(\tau_L; X),$$

$$\sigma_U|X = (1-\tau_H) * q(\tau_H; X)$$

where $q(.; X)$ is the sparsity function defined in (2). The scaling factors τ_L and $(1-\tau_H)$ guarantee the full density of birth weight integrates to 1.

This meld of semiparametric and parametric approaches has several attractive properties, the first of which is covariate-dependent threshold selection. Permitting the predictor effects to change with quantile level helps properly identify the extreme values while allowing the relationship between the predictors and the response to differ in the tail and the heart of the distribution. Second, a larger percentage of the data are used to inform our tails than in canonical extremal analysis, where observations below a threshold are often discarded. The scale parameter is the reciprocal of the conditional density at the threshold and is

influenced by values above and below the threshold. In our model moving from the tail to the middle of the distribution attenuates the effect of the observations on the tail parameters. Antecedents of this idea include (Frigessi et al., 2002) and (Behrens et al., 2004), who utilized mixture models where a light-tailed component dominated in the bulk of the distribution and a Pareto-tailed component determined the density in the tail. Finally, our hierarchical framework enables separate tails for the distribution of birth weight for each gestational age. By shrinking the predictor effects independently across basis function the degree of shrinkage can differ in the tails and the middle. The opportunity for shrinkage is especially important in the tails, where information is often limited.

In the most general case we let the shape parameter be a function of gestational age and the predictors. The quantile distribution $\mathrm{Q}^\star$ of birth weight with a Pareto tail is

$$\mathrm{Q}^\star(\tau; X_i) = \begin{cases} Q(\tau_L; X_i) - \frac{\sigma_L}{\xi_L(X_i)}\left[(\tau/\tau_L)^{-\xi_L(X_i)} - 1\right] & \tau < \tau_L \\ Q(\tau; X_i) & \tau_L \leqslant \tau \leqslant \tau_H \\ Q(\tau_U; X_i) + \frac{\sigma_U}{\xi_H(X_i)}\left[\left(\frac{1-\tau}{1-\tau_H}\right)^{-\xi_H(X_i)} - 1\right] & \tau_H < \tau. \end{cases}$$

Shape parameters can be difficult to estimate, so we favor the simplifications $\xi_L(X_i) = \xi_L$ and $\xi_U(X_i) = \xi_U$. For $\tau < \tau_L$ the effect of the j^{th} predictor is

$$\beta_j(\tau) = \beta_j(\tau_L) - \sum_{m=2}^{M} B_m(\tau)\theta_{mj}\frac{\tau_L}{\xi_L}\left[(\tau/\tau_L)^{-\xi_L} - 1\right]$$

so the predictor effects are linear across the distribution. Linearity similarly holds above the upper threshold. If the shape parameters are dependent upon the predictors then we could not write our quantile function in the form of (1).

Recall that one focus of the analysis is inference at the first percentile. We could have selected a lower threshold above 0.01 (e.g.$\tau_L = 0.05$) and extrapolated those effects to the first percentile. Instead we set the thresholds at $\tau_L = 0.01$ and $\tau_U = 0.99$. This provides the additional flexibility we desire in the tails while permitting inference across the quantiles of

interest. The effect of the parameters is continuous across the threshold, with the resulting gradual change we desire.

Allowing the shape parameter to take on positive and negative values, as in a linear link function, causes instability for shape values near zero. We assign a lognormal prior, which has positive support. For this prior the density is monotonically decreasing beyond the threshold for unbounded tails, another consideration for threshold selection. The quantile distribution $\mathrm{Q}^\star$ with exponential tails is

$$\mathrm{Q}^\star(\tau; X_i) = \begin{cases} Q(\tau_L; X_i) + \sigma_L \log(\tau/\tau_L) & \tau < \tau_L \\ Q(\tau; X_i) & \tau_L \leqslant \tau \leqslant \tau_H \\ Q(\tau_H; X_i) - \sigma_H \log\left\{(1-\tau)/(1-\tau_H)\right\} & \tau_H < \tau. \end{cases}$$

2.4 *Discrete Data*

Here we extend quantile function methodology to permit a discrete response. Treating an infant born at 36 weeks as similar to an infant born at 25 weeks yet qualitatively different from an infant born at 37 weeks is undesirable. Instead of dichotomizing by PTB, we model gestational age as interval censored values of a continuous latent process. For a reported gestational age g_i we model a continuous value $G_i \in (g_i - 0.5, g_i + 0.5)$. We find the values τ_{1i} and τ_{2i} such that $Q(\tau_{1i}; X_i) = g_i - 0.5$ and $Q(\tau_{2i}; X_i) = g_i + 0.5$. Note that τ_{1i} and τ_{2i} are the CDF evaluated at the endpoints and $\mathrm{P}(g_i - 0.5 \leqslant G_i \leqslant g_i + 0.5) = \tau_{2i} - \tau_{1i}$. This produces a likelihood of $\prod_{i=1}^{N}(\tau_{2i} - \tau_{1i})$.

Unlike models that dichotomize by PTB, this approach enables parameter effects to adapt smoothly across the distribution. Similarly modeling the effects at adjacent weeks can help compensate for measurement error in gestational age, which is generally estimated using a combination of maternal self-report of the last menstrual period and clinician judgment based on interpretation of early ultrasound examinations or other factors.

2.5 *Computation*

The computationally expensive part of our likelihood is finding the value τ_i such that $Q(\tau_i) = Y_i$. The quantile function is locally a cubic polynomial, as it is the sum of cubic polymomials, so a closed form solution exists for $\tau_i = Q^{-1}(Y_i)$. Cubic roots are numerically unstable and the solution is complicated, so we find τ_i through Newtonian recursion, where $\tau_{t+1} = \tau_t + [Q_i(\tau_t) - Y_i]/q_i(\tau_t)$ is computed until $|Q_i(\tau_t; X_i) - Y_i|$ is less than the error ϵ. For our application we chose $\epsilon = 10^{-5}$.

Previous Bayesian quantile function approaches have been computationally expensive (Reich et al., 2011) or able to preserve monotonicity for only one predictor (Tokdar and Kadane, 2011). To our knowledge this paper presents the first Bayesian quantile function model that can accommodate large sample sizes and multiple predictors. For moderate sample sizes (e.g. n = 1000) and a small number of predictors our model runs in a few minutes. To analyze the birth data set, which has 565,703 observations, we used graphics processing units (Zhou et al., 2010). The likelihood is embarrassingly parallel, so we ran our model on a graphics processing unit with 400 arithmetic cores. For a likelihood evaluation of 100,000 observations we attained a better than 20-fold improvement in computation time. With fourteen predictors and eleven separate quantile functions our final analysis ran in less than 27 hours. For the simulation study and the birth outcome analyses the posterior was sampled using the Metropolis within Gibbs algorithm, with details described in the Appendix.

3. Simulation Study

Our simulation study is designed to answer three questions about our model. First, we compare a global fit of multiple quantile functions shrunk together by a collective prior (CP), as in (3), to individual fits of the quantile functions with independent priors (IP) that do not communicate, as in (1). To test this we generate data at 5 gestational ages, each with

separate covariate effects.

$$
\begin{aligned}
Y_i|g_i &= 15 + Z(g_i) * Q_t(U_i) + X_i * 0.5 * Z(g_i) * Q_t(U_i) \\
U_i &\overset{\text{iid}}{\sim} \text{Unif}(0,1) \\
X_i &\overset{\text{iid}}{\sim} \text{Unif}(-1,1) \\
Z &\sim \text{N}_5(0,\Sigma)
\end{aligned}
$$

where Σ is the covariance matrix of a 5-dimensional first order autoregressive (AR-1) process with correlation 0.5 and unit variance. We add a scalar so the realization from this process has minimum 1 to ensure monotonicity of the quantile function. The base quantile function Q_t is the Student t quantile function with 10 degrees of freedom. We selected Student's t-distribution because our data are moderately thick-tailed and the Student t-distribution enables comparison between Pareto and exponential tails, our second factor of interest. Finally, we are interested in the loss of information due to the discretization of a continuous process. For each combination of prior and tail type we evaluate one fit with continuous response and one fit with response rounded to the nearest integer. We compare these three factors for sample sizes of n = 200 and n = 400 observations for each of the 5 levels. We ran 100 Monte Carlo replications. To complete the CP model we assign priors $\mu_{mj} \sim$ Gaussian(0,0.001), $\gamma_{mj} \sim$ Gamma(1,1), and $\rho_{mj} \sim$ Uniform(0,1). For the IP model we assign Gaussian(0,0.001) priors to the parameters for the constant basis functions and Gaussian$(\mu_j, \gamma_j^{-1}\Sigma_j)$ priors for $\theta_j = \{\theta_{2j}, ..., \theta_{Mj}\}$. The correlation matrices Σ_j are AR-1 with correlation parameter ρ_j and are designed to capture residual correlation in the regression coefficients for a predictor after shrinking to a common mean. For the IP model we assign Gaussian(0,0.001) priors to the means μ_j, Gamma(1,1) priors to the precisions γ_j and Uniform(0,1) priors to the correlations ρ_j. We fit 7 basis functions with interior knots at (0.1,0.5,0.9) and chose a lower threshold of $\tau_L = 0.01$ and an upper threshold of

$\tau_H = 0.99$. Fits with 5 basis functions were found to lack flexibility and 9 basis functions suffered from overfitting in the tails. The 7 basis function fit was preferred by log pseudo marginal likelihood (Ibrahim et al., 2001) for both the CP and IP models.

We compare our model regression estimates to the true regression effects $\beta_1(\tau) = 0.5 * Z(g_i) * Q_t(\tau)$ at the quantile levels $\tau = (0.01, 0.05, 0.10, ..., 0.90, 0.95, 0.99)$. We fit traditional frequentist quantile regression at each of the 5 gestational ages and each quantile level of interest for comparative purposes. We show summaries of ratios of Bayesian mean squared error (MSE) over frequentist MSE and coverage probabilities at the nominal 95% level averaged across week in Figure 3.

[Figure 3 about here.]

There was no difference in performance between models that received continuous data and models that received integer data so we include two lines of each combination of model type and tail type and do not visually distinguish them. This implies that there is little to no degradation in performance due to the weekly censoring for our application. For more extreme censoring, where only a few levels of the response are observed, this may not hold.

Ratios of frequentist MSE over quantile function MSE are larger than one for all models and quantile levels at both sample sizes, implying that for MSE the quantile function approach dominates the individual frequentist fits. The gains are greatest in the tails, which suffer from a paucity of information and provide the best opportunity for communicating across quantile level. The joint models outperform the individual fits with respect to MSE and coverage probability. The largest gains are in the tails and at the smaller sample size, where information is meager. The quantile functions with Pareto tails outperform the quantile functions with exponential tails at the 1st and 99th percentiles, but inference at inner quantile levels is robust to tail type. At the smaller sample size the CP model with exponential tail outperforms the IP model with Pareto tail, suggesting that for smaller sample sizes sharing

of information across models is more important than proper selection of tail type. The CP and the Pareto models were selected as best by log pseudo marginal likelihood (LPML).

To test our model for distributions with bounded support we generated responses $Y_i|g_i = 20 + 20 * Q_1(U_i) + 10 * (Z(g_i) + X_i) * Z(g_i) * Q_2(U_i)$ where $Q_1(U_i)$ is the quantile function of a Beta(1,5) random variable and $Q_2(U_i)$ is the quantile function of a Beta(5,1) random variable. Results are presented in the online appendix. Bayesian methods again dominated the frequentist method with respect to MSE. Bayesian methods achieved near nominal coverage probability for all quantile levels for n = 400. Coverage probability in the tails was slightly higher for the models with exponential tails.

4. Birth Outcome Analyses

4.1 *Data Description*

We incorporate several sources of data in our analyses. The birth data consist of live birth certificate records in Texas from 2002 - 2004. Personal characteristics consisted of infant sex, maternal parity (binary for having given birth previously), maternal age and paternal age (less than 40 or 40 and above), maternal education and paternal education (did not finish high school, finished high school or finished some education after high school), and maternal ethnicity (white non-Hispanic, black non-Hispanic, Hispanic and other). All personal characteristics were treated as categorical variables. Data from pregnancies that were missing key covariates, or ended with spontaneous fetal death or induced termination were discarded. We analyze viable births, those occurring in weeks 25-42 (Morgan et al., 2008), leaving 565,703 live births. Sample size varied considerably by region and gestational age. Public Health Region (PHR) 9 in midwest Texas had the smallest number of infants in our sample (9936), while PHR 3, which includes Dallas-Fort Worth, had the most (155,958.) Only 467 infants were born at 26 weeks while 157,689 were born at 39 weeks.

Pollution data come from the Environmental Protection Agency's downscaler model (Berrocal et al., 2010). The downscaler model integrates monitor measurements from the Environmental Protection Agency (EPA) Air Quality System network with deterministic numerical output data from the EPA's Community Multiscale Air Quality (Byun and Ching, 1999) model. The downscaler model assumes a linear relationship between the biased numerical output data and the spatio-temporally sparse monitor data. We mapped each maternal address to the nearest 12 kilometer by 12 kilometer grid cell centroid and summed these values for the first and second trimesters.

4.2 *Gestational Age - Personal Characteristic Results*

We examine the effects of the predictors for quantile levels of
$\tau = (0.01, 0.05, 0.10, ..., 0.90, 0.95, 0.99)$ for both responses. We used Gaussian(0,0.001) priors for the means and Gamma(1,1) priors for the precisions. For the range parameters ϕ_{mj} we implemented Uniform(0,r) priors, where r is half of the maximum distance between Texas PHR centroids. We chose a lower threshold of 0.01 and an upper threshold of 0.99. For 5 and 7 basis functions we fit exponential tails, Pareto tails with common shape across PHR, and Pareto tails with shapes varying across PHR. For the shape varying across PHR for both upper and lower tails we assigned on the log scale a multivariate normal prior with mean ξ_μ and precision ξ_γ and spatially correlated the tails with an exponential correlation structure. We assigned ξ_μ a normal (-0.4, .16) prior and ξ_γ a Gamma(6,1) prior for upper and lower tails. For comparison we fit separate quantile functions independently across PHR. The fit with 7 basis functions and shape varying across PHR had the highest LPML, as can be seen in Table 1.

[Table 1 about here.]

The independent fits across PHR that had individual shape parameters had higher LPML than the collective models with a common shape parameter, indicating the need for the shape

parameter to vary across region. With most of the births occuring during weeks 31-42, we did not feel comfortable going above 7 basis functions for only 12 levels of gestational age.

The effects of black non-Hispanic mothers relative to white non-Hispanic mothers on gestational age are presented in Figure 4.

[Figure 4 about here.]

Black non-Hispanic maternal ethnicity is associated with a more than 2 week decrease in gestational age at the first percentile. The effect diminishes for large quantiles and there is no significant difference for upper quantiles. This illustrates the ability of quantile function modeling to caputure different covariate effects on different aspects of the distribution. Other practically significant predictors included maternal parity and maternal education above high school, which were positively associated with gestational age, and maternal age above 40, which was negatively associated with gestational age in the lower tail. All posterior quantile plots for gestational age are available in the online Appendix.

4.3 *Birth Weight Personal Characteristic Results*

Allowing the effects of the predictors on birth weight to vary by region and gestational age would have been computationally expensive for 14 predictors, so the predictor effects on birth weight are constant across region. We used the same thresholds, mean, precision and tail priors that were used in the gestational age analysis. We selected Uniform(0,1) priors for the correlation parameters ρ_{mj}. We implemented fits of 7,9 and 11 basis functions with exponential tails, Pareto tails with shape parameter shared across gestational age and Pareto tails correlated across gestational age through an AR-1 prior on the log scale. Table 1 shows that the fit with 9 basis functions with shape parameter shared across gestational age was preferred by LPML.

Infants born to black non-Hispanic mothers weigh between 20-190 grams less than infants born to white non-Hispanic mothers. The effect is larger for higher gestational ages and

change little with quantile level. Infants from black non-Hispanic mothers are at high risk for both PTB and SGA, but the relationship between the birth outcomes and risk of morbidity and mortality is subtle. While PTB is associated with negative health outcomes, previous research has shown that certain groups of PTB black non-Hispanic infants have better survival rates than PTB white non-Hispanic infants (Schieve and Handler, 1996). LBW black non-Hispanic infants also have lower mortality rates than their white non-Hispanic counterparts (Alexander et al., 2003). Given health outcome data, the methods presented in this paper could be used to better understand how the predictors affect morbidity and mortality indirectly through the birth outcomes.

While most of the effects of the predictors on gestational age are not practically significant, almost all of the personal characteristics substantially affect birth weight. The uncertainty in the posterior distribution of the regression parameters is smallest for gestational ages with the largest sample sizes, which occur during weeks 37-40. Uncertainty is generally higher in the upper tail than the lower tail. As can be seen in Figure 1, the distribution of birth weight is right-skewed for small gestational ages but becomes more symmetric as gestational age increases. This is possibly the reason posterior variances for birth weight effects are higher in the upper tail than the lower tail for smaller gestational ages. We present the personal characteristic effects in the online Appendix.

4.4 *Ozone Results*

For gestational age we permit the ozone effects to vary by region. Ozone levels are higher in cities, so allowing the effects to change regionally allows ozone effects to differ for urban and rural areas. The distribution of unmeasured potential confounders may vary regionally, so allowing the ozone effects to vary spatially may help adjust for this.

As can be seen in Figure 5, second trimester ozone is negatively associated with gestational age in Texas Public Health Regions 4, 5, 6 and 11, located in east and south Texas.

[Figure 5 about here.]

The largest effect in absolute magnitude was 3-4 days for second trimester ozone at the 40^{th} percentile in Region 11. First trimester ozone had negative effects in Public Health Regions 5 and 11. Public Health Regions 2, 3, 4, 6 and 9 showed slight negative associations in the lower tail with first trimester ozone. Previous research using birth records from Harris county, whose county seat is Houston and is located in Region 6, found evidence that high levels of first trimester ozone increased the probability of PTB (Warren et al., 2012). While almost all of the ozone effects are strongest in the lower tail, in Public Health Region 11 the strongest effects are located around the 40^{th} percentile for both first and second trimester ozone. In our sample 89% of Public Health Region 11 is Hispanic, compared to 49% for Texas. Public Health Region 11 has many migrant workers, high levels of poverty and high prevalence of birth defects (Hendricks et al., 1999), so it is unsurprising we find that the nature of the relationship between gestational age and ozone differs here from the rest of Texas.

While Frequentist and Bayesian models found second trimester ozone to be associated with 10-15 gram decreases in birth weight for gestational ages 34-42 weeks, Figure 6 shows the large reduction in variance that is possible by borrowing information across gestational age.

[Figure 6 about here.]

By modeling the full quantile function, the individual fits are smoother and generally have smaller variance than frequentist quantile regression. The difference in the variances is largest for gestational ages with smaller sample sizes (e.g. 35 weeks). At 38 and 39 weeks of gestational age, where the sample size is largest and there is less need to borrow information from neighboring gestational ages, the Bayesian regression parameters change continuously and slowly across quantile level. In contrast, the frequentist effects appear choppy and have larger standard errors because they do not borrow information across quantile level. Second trimester ozone results were statistically significant for gestational ages of 32-41 weeks.

We did not find a statistically significant relationship between first trimester ozone and birth weight. Blood exchange between the mother and the placenta increases at the beginning of the second trimester. The negative association between fetal growth and large levels of ozone in the second trimester could be due to impairment of uteroplacental blood flow, which has been shown to be affected by maternal smoking (Mochizuki et al., 1984).

5. Conclusions

In this paper we have presented a novel class of hierarchical quantile function models that retains the interpretability of quantile regression and some of the malleability of density estimation. Our hierarchical framework permits flexible tail inference and can model discrete response. While our application consists only of interval censored data, it is easily extended to data that are left or right censored or a mixture of continuous and discrete responses. Our discrete quantile model could be a viable alternative to other applications, such as Likert scale response.

We conducted a simulation study that demonstrated substantial reductions in MSE for our approach relative to canonical quantile regression for both thickly-tailed and bounded distributions. We found that the lower tails of the distributions of gestational age and birth weight for infants born to African American mothers are much lower than the lower tails for infants born to white non-Hispanic mothers. Useful extension of our methods include variable selection for a large number of predictors and incorporating subject-specific random effects.

Acknowledgements

The authors are grateful for support from NSF grant DMS-1107046 and NIH grant SR01ES014843. This work is partially supported through cooperative agreement U01DD000494 between the Center for Disease Control and the Texas Department of State Health Services.

References

Alexander, G. R., Kogan, M., Bader, D., Carlo, W., Allen, M., and Mor, J. (2003). US birth weight/gestational age-specific neonatal mortality: 1995-1997 rates for whites, hispanics, and blacks. *Pediatrics* **111,** e61 – e66.

Barker, D. J. (2006). Adult consequences of fetal growth restriction. *Clinical obstetrics and gynecology* **49,** 270–283.

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling* **4,** 227–244.

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010). A bivariate space-time downscaler under space and time misalignment. *The annals of applied statistics* **4,** 1942.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika* **97,** 825–838.

Burgette, L. F. and Reiter, J. P. (2012). Modeling adverse birth outcomes via confirmatory factor quantile regression. *Biometrics* **68,** 92–100.

Byun, D. W. and Ching, J. (1999). *Science algorithms of the EPA Models-3 community multiscale air quality (CMAQ) modeling system*. US Environmental Protection Agency, Office of Research and Development Washington, DC, USA.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

DeVore, R. A. (1977). Monotone approximation by splines. *SIAM Journal on Mathematical Analysis* **8,** 891–905.

Dunson, D. B., Herring, A. H., and Siega-Riz, A. M. (2008). Bayesian inference on changes in response densities over predictor clusters. *Journal of the American Statistical Association* **103,** 1508–1517.

Frigessi, A., Haug, O., and Rue, H. (2002). A dynamic mixture model for unsupervised tail

estimation without threshold selection. *Extremes* **5,** 219–235.

Gardosi, J., Mongelli, M., Wilcox, M., and Chang, A. (1995). An adjustable fetal weight standard. *Ultrasound in Obstetrics & Gynecology* **6,** 168–174.

Garite, T. J., Clark, R., and Thorp, J. A. (2004). Intrauterine growth restriction increases morbidity and mortality among premature neonates. *American journal of obstetrics and gynecology* **191,** 481–487.

Hendricks, K. A., Simpson, J. S., and Larsen, R. D. (1999). Neural tube defects along the texas-mexico border, 19931995. *American Journal of Epidemiology* **149,** 1119–1127.

Honein, M. A., Kirby, R. S., Meyer, R. E., Xing, J., Skerrette, N. I., Yuskiv, N., Marengo, L., Petrini, J. R., Davidoff, M. J., Mai, C. T., et al. (2009). The association between major birth defects and preterm birth. *Maternal and child health journal* **13,** 164–175.

Ibrahim, J. G., Chen, M.-H., and Debajyoti, S. (2001). *Bayeisan Survival Analysis*. Springer, New York.

Kammann, E. E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **52,** 1–18.

Katz, J., Lee, A. C., Kozuki, N., Lawn, J. E., Cousens, S., Blencowe, H., Ezzati, M., Bhutta, Z. A., Marchant, T., Willey, B. A., et al. (2013). Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: a pooled country analysis. *The Lancet* .

Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives* **15,** 143–156.

Machado, J. A. F. and Silva, J. S. (2005). Quantiles for counts. *Journal of the American Statistical Association* **100,** 1226–1237.

Mochizuki, M., Maruo, T., Masuko, K., Ohtsu, T., et al. (1984). Effects of smoking on

fetoplacental-maternal system during pregnancy. *American journal of obstetrics and gynecology* **149,** 413–420.

Morgan, M. A., Goldenberg, R. L., and Schulkin, J. (2008). Obstetrician-gynecologists' practices regarding preterm birth at the limit of viability. *Journal of Maternal-Fetal and Neonatal Medicine* **21,** 115–121.

Narchi, H., Skinner, A., and Williams, B. (2010). Small for gestational age neonates-are we missing some by only using standard population growth standards and does it matter? *Journal of Maternal-Fetal and Neonatal Medicine* **23,** 48–54.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association* **74,** 105–121.

Pulver, L. S., Guest-Warnick, G., Stoddard, G. J., Byington, C. L., and Young, P. C. (2009). Weight for gestational age affects the mortality of late preterm infants. *Pediatrics* **123,** e1072–e1077.

Ramsay, J. (1988). Monotone regression splines in action. *Statistical Science* pages 425–441.

Reich, B. J., Cooley, D., Foley, K. M., Napelenok, S., and Shaby, B. A. (2011). Extreme value analysis for evaluating ozone control strategies.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association* **106,**.

Reich, B. J. and Smith, L. B. (2013). Bayesian quantile regression for censored data. *Biometrics* .

Rogers, J. F. and Dunlop, A. L. (2006). Air pollution and very low birth weight infants: A target population? *Pediatrics* **118,** 156–164.

Schieve, L. A. and Handler, A. (1996). Preterm delivery and perinatal death among black and white infants in a chicago-area perinatal registry. *Obstetrics & Gynecology* **88,** 356–363.

Schumaker, S. (1981). *Spline Functions: Basic Theory.* John Wiley & Sons, Inc., United

States of America.

Šrám, R. J., Binková, B., Dejmek, J., and Bobak, M. (2005). Ambient air pollution and pregnancy outcomes: a review of the literature. *Environmental health perspectives* **113,** 375.

Tokdar, S. and Kadane, J. B. (2011). Simultaneous linear quantile regression: A semiparametric bayesian approach. *Bayesian Analysis* **6,** 1–22.

Tukey, J. W. (1965). Which part of the sample contains the information? *Proceedings of the National Academy of Sciences of the United States of America* **53,** 127.

Wang, H. and Tsai, C.-L. (2009). Tail index regression. *Journal of the American Statistical Association* **104,** 1233–1240.

Wang, H. J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association* **107,** 1453–1464.

Warren, J., Fuentes, M., Herring, A., and Langlois, P. (2012). Spatial-temporal modeling of the association between air pollution exposure and preterm birth: Identifying critical windows of exposure. *Biometrics* **68,** 1157–1167.

Wilcox, A. J. (2001). On the importance-and the unimportance-of birthweight. *International journal of epidemiology* **30,** 1233–1241.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54,** 437 – 447.

Zhou, H., Lange, K., and Suchard, M. A. (2010). Graphics processing units and high-dimensional optimization. *Statistical Science* **25,**.

Zhou, J., Chang, H. H., and Fuentes, M. (2012). Estimating the health impact of climate change with calibrated climate model output. *Journal of agricultural, biological, and environmental statistics* **17,** 377–394.

Received October 2013.

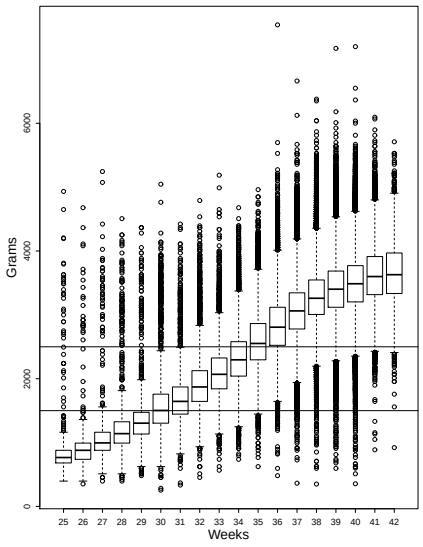

Figure 1. Boxplots of birth weight by week of gestational age

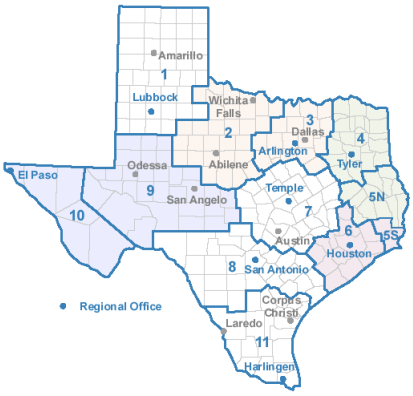

Figure 2. Texas Public Health Regions

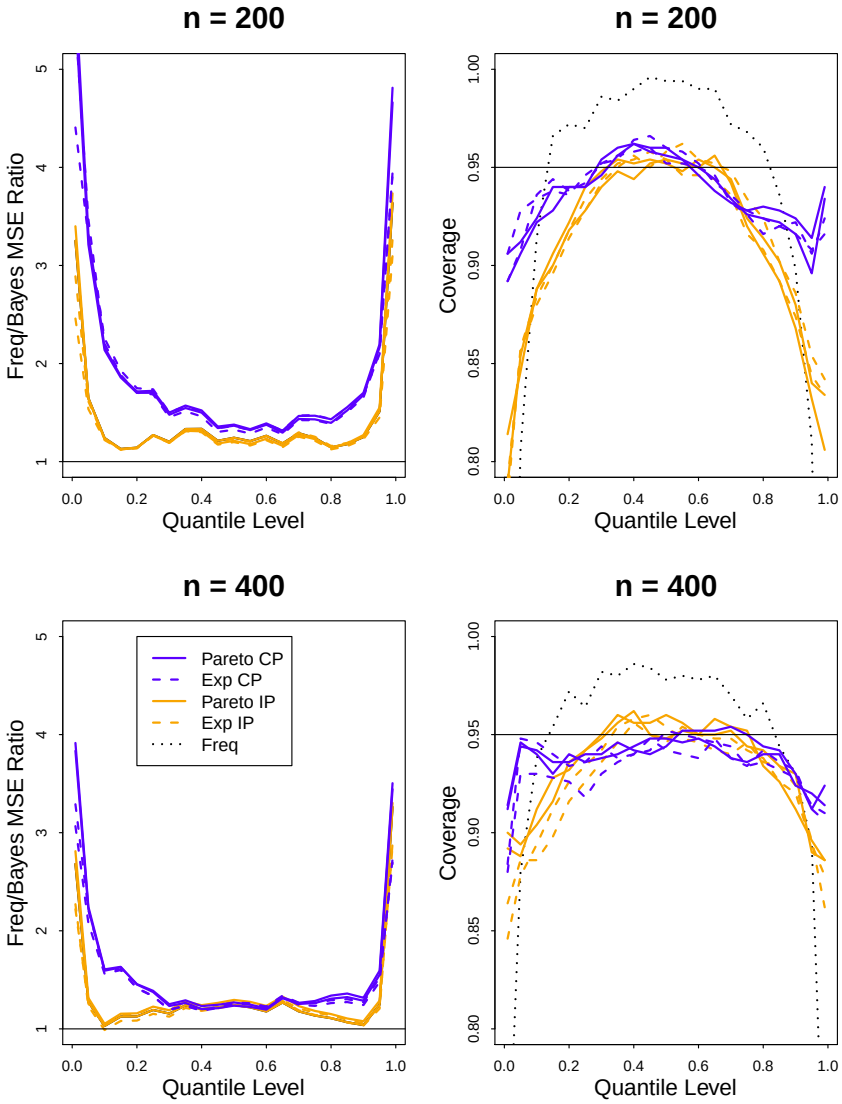

Figure 3. MSE Ratios of quantile function estimator over frequentist quantile estimator and coverage probabilities. The maximum Monte Carlo standard error of MSE was 0.35 for Bayesian estimators and 1.2 for the frequentist estimator. The maximum Monte Carlo standard error for the MSE ratio was 0.53 at $\tau = 0.01$ and 0.13 for other quantile levels.

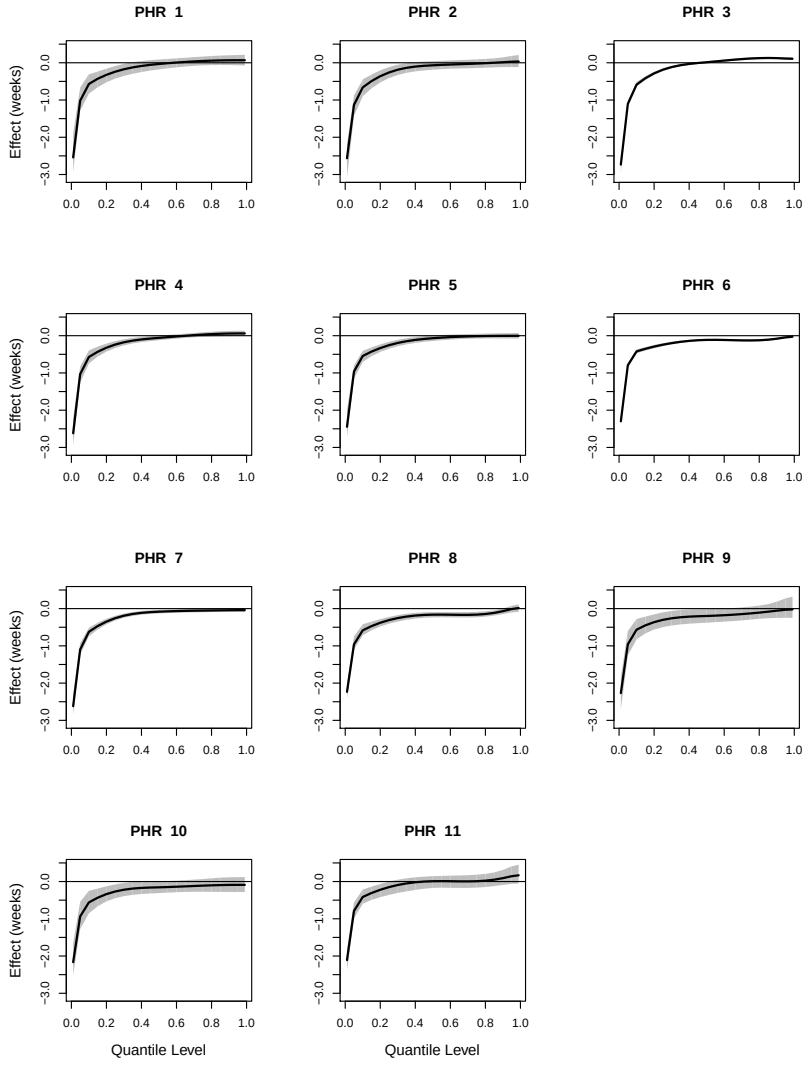

Figure 4. 95% credible limits for the posterior distribution of the difference in gestational age between black non-Hispanic and white non-Hispanic mothers by Public Health Region (PHR).

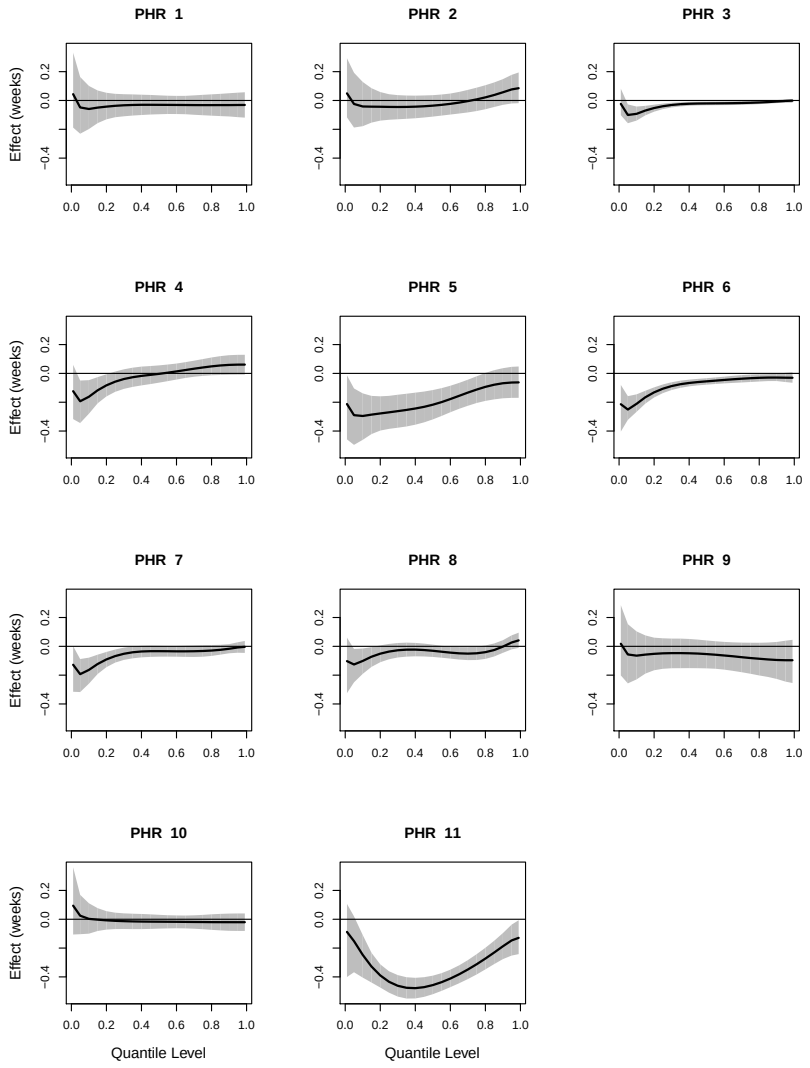

Figure 5. 95% credible limits for the posterior distribution of the effect of a one-unit increase in second trimester ozone exposure on gestational age by Public Health Region. All ozone values were linearly transformed into [-1,1], so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high.

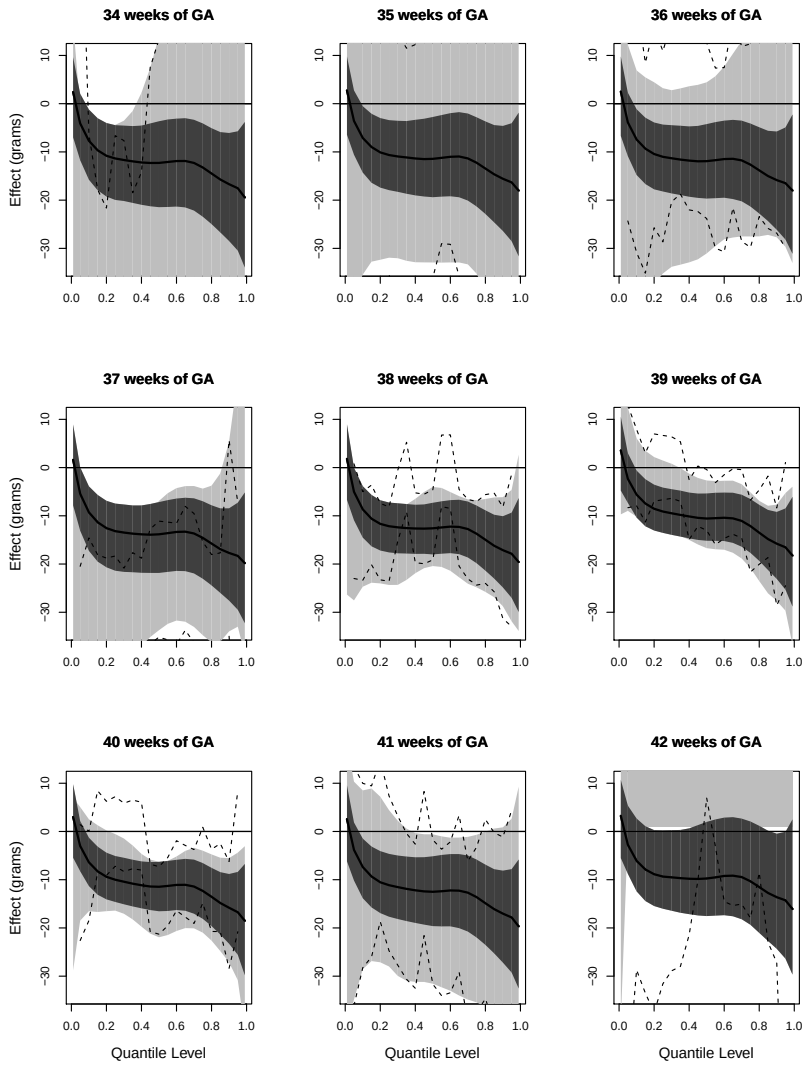

Figure 6. 95% credible limits for the posterior distribution of the effect of a one-unit increase in second trimester ozone exposure on birth weight for gestational age of 34-42 weeks. All ozone values were linearly transformed into [-1,1], so a one-unit increase can be roughly thought of as an increase from low levels to middle levels of exposure, or middle to high. Light gray regions correspond to posterior credible sets for individual fits at each gestational age while dark gray regions correspond to the collective fit across gestational age. Dashed lines indicate limits of 95% frequentist confidence intervals.

Table 1

Log pseudo marginal likelihood (LPML) of model fits for the birth outcomes, with higher values corresponding to better fits. Minimum values for gestational age (-1,077,185) and birth weight (-4,094,739) were subtracted from the values shown below for clarity. Model types include exponential tail (Exp), Pareto tail with shape constant for all Public Health Regions or gestational ages (Par), shape changing across Public Health Region or gestational age (Change), and independent models for each Public Health Region and gestational age (Ind). Bolded values signify the best fit.

	Number of Basis Functions	Model Type	LPML
Gestational age	5	Exp	0
	5	Par	60,965
	5	Change	60,908
	7	Exp	91,085
	7	Par	91,164
	7	Ind	91,281
	7	Change	**91,388**
Birth weight	7	Exp	227
	7	Par	1663
	7	Change	1665
	9	Exp	764
	9	Par	**2411**
	9	Ind	0
	9	Change	2373
	11	Exp	838
	11	Par	2261
	11	Change	2398